Genetic diversity and population structure among oat cultivars and landraces

Gracia Montilla-Bascón[1], Javier Sánchez-Martín[1], Nicolas Rispail[1], Diego Rubiales[1], Luis AJ Mur[2], Tim Langdon[2], Irene Griffiths[2], Catherine Howarth[2] and Elena Prats[1].

[1]CSIC, Institute for Sustainable Agriculture, Apdo. 4084, E-14080 Córdoba, Spain.

[2]Institute of Biological, Environmental and Rural Sciences, University of Aberystwyth, UK

*To whom correspondence should be addressed:

Elena Prats

Institute for Sustainable Agriculture, CSIC, Apdo. 4084, E-14080 Córdoba, Spain.

E-mail: elena.prats@ias.csic.es

Tel: +34 957499291

Fax: +34 957499252

Abstract

In this study genetic diversity among 177 oat (*Avena sativa* L.) accessions including both, white and red oats landraces and 36 commercial cultivars, was studied for simple sequence repeat (SSR) loci. Thirty one genomic and expressed sequence tags (EST) derived primer pairs were selected according high polymorphism from an initial 66 SSR batch. Markers revealed a high level of polymorphism, detecting a total of 454 alleles. The average gene diversity for the whole sample was 0.29. Genetic similarity, calculated using the Dice coefficient, was used for cluster analysis and principal component analysis was also applied. In addition, population structure using a Bayesian clustering approach identified discrete subpopulation based on allele frequency and showed similar clustering of oat genotypes in four groups. Accessions could be classified into four main clusters that clearly separated the commercial cultivars, the red oats landraces and two clusters of white oat landraces. Cultivars showed less diversity than the landraces indicating a reduction of genetic diversity during breeding, whereas white oat landraces showed higher diversity than red ones. The average polymorphic information content of 0.80 for the SSR loci indicated the usefulness of many of the SSR for genotype identification. In particular two markers, MAMA5 and AM04, with a total of 50 alleles and a high discrimination power (>0.90) were sufficient to discriminate among all commercial cultivars studied highlighting their potential use for variety identification.

Keywords: *A. sativa,* genetic diversity, oat, simple sequence repeat (SSR),

Introduction

Oat is a cereal crop of Mediterranean origin (Stevens et al. 2004). *Avena sativa* L. is the main cultivated oat including the white and red oats. White oats are preferred for milling and are used for human food and for fodder, especially poultry and horses. Red oats (formerly known as *A. byzantina* K. Koch) are preferred for hay (Stevens et al. 2004).

During the 20th century, landraces have increasingly been replaced by modern cultivars, resulting in significant reduction in genetic diversity (Warburton et al. 2008; Reif et al. 2005; Roussel et al. 2004) and contributing to the stability in genetic diversity of wheat, barley and maize cultivars in recent years (Christiansen et al. 2002; Donini et al. 2000; Koebner et al. 2003). Thus, the loss of genetic diversity has become an important problem both in natural plant populations and in important crop species. This loss led to calls for the genetic conservation of crop germplasm (Frankel and Bennett 1970). Current molecular characterization of *ex situ* plant germplasm has placed more emphasis on cultivated gene pools and less on exotic gene pools representing wild relative species (Karp 2002). Although these modern cultivars may be higher yielding under high input systems, landraces have considerable potential for use in improving disease and abiotic stress tolerance. Transfer of beneficial traits from landraces is relatively straight-forward in that there is no barrier to crossing as there can be with the use of crop wild relatives. Thus, several studies suggest that landraces may be a good source of new allelic diversity for breeding programmes. However, better characterization of exotic germplasm is needed to facilitate its use in plant breeding and

in research (Hawkes 1990; Jellen and Leggett 2006), so as to the introgression of exotic germplasm into a plant breeding program.

Genetic diversity studies, assessed by various tools including DNA markers, provide important information both for genetic conservation and for use in efficiently breeding new commercial varieties. To date, genetic studies in hexaploid oat has been more difficult than in other species, mainly due to large genome size (Bennett and Smith 1976) and polyploidy causing inherent complexities for mapping including a large numbers of linkage groups, detection of multiple loci by a single probe, and co-migration of fragments from different loci that can impede interpretation of allelic relationships and genetic analyses (Iannucci et al. 2011). Amplified fragment length polymorphisms (AFLPs) (Achleitner et al. 2008; Fu et al. 2005; Fu and Williams 2008), random amplified polymorphic DNA (RAPDs) (Baohong et al. 2003; Paczos-Grzeda 2004) and microsatellites have been previously used in oat for assessment of genetic diversity. In particular, microsatellites have been used to dissect genetic diversity in several *Avena* spp. (Li et al. 2000; Li et al. 2007) and to examine allelic diversity changes over 100 years of oat breeding in both Nordic countries (Nersting et al. 2006), Canada (Fu et al. 2007) and North Europe (He and Bjornstad 2012). Association of genetic markers with regions of the genome controlling different traits would enable efficient and precise transfer of useful alleles from landraces to modern cultivars whilst minimizing linkage drag of non- beneficial alleles.

To date identification of oat cultivars has relied on morphological and phenological characteristics that may be influenced by environmental factors and require trained staff and large-scale growth experiments of mature plants under uniform

conditions for evaluation. In addition some cultivars are morphologically similar, making difficult to distinguish between them visually. SSR profiles can be used as a DNA fingerprint for registered cultivars to avoid redundancy of identical cultivars as well as to protect breeders´ rights.

Here, we studied the genetic diversity of 141 white and red oat landraces together with 36 currently grown oat cultivars for potential use in breeding programmes. Furthermore we tested the potential of SSRs for molecular identification of the oat cultivars studied.

Materials and methods

Plant material

A germplasm collection of landraces consisting of 141 *A. sativa* accessions (110 white and 31 red oats) originally collected from 1944 to 1997 in southern Spain, when they were used locally in agriculture (Online Resource 1), was provided by the "Plant Genetic Resources Center" (CRF-INIA, Madrid, Spain). In addition, 36 commercial cultivars were supplied by the Andalusian Network of Agriculture Experimentation (RAEA) selected for their adaptation to southern Spain agroecological conditions. For simplicity, germplasm bank codes were substituted for the codes included in Supplementary Table 1 (Sánchez-Martín et al. 2011a). White oat cultivars studied were Ac1, Acebeda, Adamo, Aintree, Alcudia, Anchuela, Araceli, Brawi, Caleche, Canelle, Chambord, Chapline, Charming, Cobeña, Condor, Cory, Edelprinz, Flega, Fringante, Fuwi, Hammel, Kankan, Kantora, Karmela, Cassandra, Kazmina, Mirabel, Mojacar,

Norlys, Orblanche, Pallini, Patones, Prevision, Primula, and Rapidena. In addition, the *A. strigosa* cultivar Saia was included for comparison.

Seedlings were grown in 0.5 L pots filled with peat:sand (3:1) in a growth chamber at 20 ºC, 65% relative humidity and under 12 h dark/12 h light with 150 µmol m^{-2} sec^{-1} photon flux density supplied by high-output white fluorescent tubes. Sites and year of landraces sampling together with other characteristics of the site are recorded in Online Resource 1 and year of registration, origin, and genealogy if known of cultivars are recorded in Online Resource 2.

DNA extraction, SSR markers and PCR procedure

Ten leaves from 12-day-old seedlings were harvested and DNA extracted according to the CTAB protocol (Murray and Thompson 1980). Sixty-six SSR primer pairs derived from genomic and EST libraries of oats and barley were selected from previous reports to test for polymorphism (Becher 2007; Jannink and Gardner 2005; Li et al. 2000; Liu et al. 1996; Pal et al. 2002; Wight et al. 2010). In addition, 3 SSR were developed from EST sequence information, from *Avena barbata* and *Festuca* by using the following primers: Barb2-40, (5'-CCATCTCAACCTTTGCTTCTCTCCT-3' and 5'-GTTCTTGAGCTCCTTGACCTTGAGC-3'); Barb4-10, (5'-GCTGAGCAATCTCA TCAGCTCAACT-3' and 5'-GAGGTGATCCGAGCTTACTTCATCA-3'); Fesc12, (5'-GTCGCCGGAGAAGAGAAGAG-3' and 5'-AACGCTAGCCGTGATGACTT-3'). Following preliminary assays in a subset of 46 samples a final set of 31 primer pairs (Table 1) were chosen because of their consistency in amplification and polymorphism in our oat genotypes and/or because they had been mapped in a mapping population

developed from two winter oat cultivars Buffalo and Tardis (data not shown) and displayed reasonable genome coverage.

Amplification reactions were set up for 40 cycles with an initial denaturing step of 10 min at 95 °C. Each cycle consisted of denaturation at 94 °C for 1 min, followed by primer-specific annealing for 1 min (temperature specified in Table 1) and extension at 72 °C for 1 min. After 35 cycles, there was a final extension step of 10 min at 72 °C. The 10 μL reaction mix comprised 6.05 μL sterile-distilled water, 0.05 μL Taq polymerase (Roche Applied Sciences, Mannheim, Germany), 1 μL of 10x PCR buffer with $MgCl_2$, 1.3 μL of dNTP (5 mM equimolar solution of each dATP, dCTP, dGTP and dTTP), 0.3 μL each of forward and reverse primers (10 μM solution) and 1 μL of template DNA (30 ng μL^{-1}). Reactions were stopped with 95% formamide loading dye. Amplification products from markers AME097, AME105, AME168, AME176, AME192, BarbSSR_2-40, and BarbSSR_4-10 were separated on 4.5% polyacrylamide denaturing gel (BIO-RAD, California, USA, Sequi-GenGT, 38 × 50 cm) using 73-well comb and visualized by silver staining (Promega Silver Sequencing system, Wisconsin, USA). Relative movement of different amplicons and standard molecular-weight marker was used to estimate the sizes of amplified fragments using regression. The remained markers were run on the ABI 3137 capillary sequencer. PCR, using AmpliTag gold, was conducted as for polyacrylamide gel analysis except that one primer was labeled with a fluorescent dye and the concentration of DNA was 20 ng μL^{-1}. The size standard Genescan500 LIZ^{TM} (orange) was included with each sample and used to determine the sizes of the PCR products detected. All primers and the size standards were supplied by Applied Biosystems (ABI). Data were analysed using Genemapper (ABI). Presence or absence of each amplified band was scored as 1 and 0, respectively,

for all markers to generate a binary data matrix. The genetic diversity of each microsatellite locus was assessed by calculating the frequency of the microsatellite alleles based on polymorphic information content (PIC) following (Botstein et al. 1980) using the equation:

PIC= 1-J=1- $\sum_{J=1}^{n} P^2_{ij}$

where *P*ij is the frequency of the j^{th} allele for i^{th} marker. Estimates of genetic similarity (GS) were calculated for all possible pairs of genotypes according to Dice similarity coefficient (Nei and Li 1979). In addition, frequencies of incidence of all polymorphic alleles for each SSR marker were calculated and used for determining statistical parameters. Confusion probability (Cj) and discriminating power (Dj) of each marker were estimated according to Tessier et al. (1999). Cluster analysis based on unweighted pair-group method with arithmetic average (UPGMA) was performed on matrix of GS estimates using GenStat 7^{th} Edition and a dendrogram constructed. The correlation coefficient between the similarity matrix and the cophenetic values matrix was computed to test the goodness of fit of the cluster analysis. NTSYS-pc 2.02j software (Biostatistics Inc., USA, Rohlf 1998) was used for these statistical analyses.

Population structure and percentages of admixture

Population structure was inferred by the software STRUCTURE 2.3.3. We set most parameters to their default values as advised in the user's manual (Pritchard and Wen 2003). Specifically, we chose the admixture model and the option of correlated allele frequencies between populations, as this configuration is considered best in cases of subtle population structure (Falush et al. 2003). Similarly, we let the degree of admixture alpha be inferred from the data. Each simulation included 10,000 burn-in and

100,000 iterations. Longer burn-in or MCMC did not change significantly the results. Ten independent simulations per k value were run and the mean estimate across runs of the log posterior probability of the data for a given k, Pr(X|k), called L(k) were plotted for each k category on a graph to determine the k value of the population as the value of k for which the distribution of L(k) plateaus or continues to increase, but much more slowly. Because this point is known to be difficult to determine, we also used Δk, an ad hoc quantity proposed by Evanno et al. (2005) related to the second order rates of change of the likelihood function with respect to k that is supposed to show a clear Δk at true value of k. The percentages of admixture of each accession (Q matrix) given by the software were used as cofactors in the association analyses. For trait analyses per subpopulation, an accession was assigned to a subpopulation when it showed more than 80% membership in this subpopulation (Figueiredo et al. 2010).

Results

The thirty-one primer pairs used to characterize and evaluate the genetic diversity in the landraces and commercial varieties of the oat collection showed a high level of polymorphism, displaying a total of 454 alleles. The total number of alleles per marker ranged from 3 for AME168 to 42 for MAMA05 with a mean of 14.65 (Table 1). PIC varied from 0.46 (AME168 and AM112) to 0.96 (AM04) with a mean of 0.80. Based on PIC values obtained, most SSRs, with the exception of AM112, AME168, AME176, BarbSSR_2-40 and BarbSSR_4-10, were considered informative markers (PIC>0.7), indicating the potential use of this set of SSR markers for cultivar identification (Table 1). Allelic frequencies observed ranged from 0.001 to 0.69 with a mean of 0.062. One hundred thirty eight alleles out of 454 detected were classified as “rare” due to their low

frequency (<0.03), 194 were classified as “common”, with frequencies between 0.03 and 0.2, and 122 were classified as “more frequent” with frequencies >0.2 (Table 1). Rare and common alleles were detected at 26 and 27 SSR loci studied, respectively. Rare alleles per locus ranged from 0 to 16 (MAMA05) whereas the number of common alleles per locus ranged from 0 to 28 (AM04) and the more frequent from 0 to 9 (AM30) (Table 1). High values of discriminating power (Dj≥0.81) and PIC≥0.81, and low values of confusion probability (Cj≤0.19) were obtained for 20 of the markers evaluated (64.5%) (Table 1).

From the dendrogram generated, the 177 accessions could be classified into four main clusters that clearly separated the commercial cultivars, the red oats landraces and two clusters of white oat landraces (Fig 1). The *A. strigosa* genotype, Saia, did not cluster with the *A. sativa* entries. Genetic similarity estimates calculated among the oat collection varied from 0.16 to a maximum of 0.99 (between Gen130 and Gen131) with a mean similarity of 0.29 (Figure 1). Cluster 1 included most of the commercial cultivars together with the landraces Gen141 and Gen17. Cultivars with the highest genetic similarity (0.88) were Chapline and Cobeña. Cluster 2 comprised the red oat landraces along with the commercial red oat cultivars Cassandra and Prevision. In addition four landraces, Gen139, Gen64, Gen27, Gen106 which are all described as white oats were included in this cluster. The red oat landrace Gen84 however grouped in cluster 3. The third and fourth cluster included most of the white oat landraces. The third cluster contained 51 genotypes with Gen5 and Gen 13 being the most related landraces with a GS of 0.84. The fourth cluster was the largest with 53 genotypes. Genetic similarity within each cluster was similar with values of 0.50, 0.55, 0.50 and 0.54 for cluster 1, 2, 3, and 4 respectively but when comparing the white oats,

commercial cultivars had a slightly higher GS compared with landraces. Most clusters showed particular alleles characteristic for that group. For example, most genotypes of cluster 1 had the 209 bp allele from AM01 marker, all genotypes of cluster two had the 137 and 368 bp alleles from AME097 and MAMA3 marker, all genotypes from cluster 3 and 4 except one had the 232 pb allele amplified with AM112 marker, but only genotypes from cluster 3 had the 161 pb allele amplified with AM87. Overall, comparison of the white oats revealed 22 unique alleles belonging to commercial cultivars and 129 to landraces. In addition 11 alleles were assigned exclusively to the red oat landraces. Analysis of the allelic frequency obtained for each marker showed that often clusters 2, 3, or 4 associated with landraces had a significantly higher allelic frequency than cluster 1 (Table 2). In particular markers AME105 and MAMA 9 showed significantly higher allelic frequences in clusters 2, 3 and 4 compared with cluster 1. Only for marker MAMA11 was the allelic frequency of cluster 1 higher than that of clusters 2, 3, and 4 (Table 2). Comparison between the cophenetic matrix and the similarity matrix was significant with r^2=0.876 ($p<0.01$) indicating a high goodness of fit between both matrices.

Interestingly, when assessing independently the oat cultivars, a subset of 11 markers (HvXan, AM01, AM30, AM14, AM87, AM07, MAMA08, AM42, MAMA5, AM04, FESC12) showed a high polymorphism with 167 alleles and high PIC (>0.80). This subset of SSRs were able to group the white oat cultivars from cluster 1 showing similar genetic relationship among them as obtained when assessed with the complete SSR set. Furthermore, among these SSRs, two of them MAMA5, with 23 alleles and a discriminating power of 0.90, and AM04, with 27 alleles and a discrimination power of 0.94, were sufficient to discriminate between all cultivars.

Structure of the population

The criteria used to define the number of subpopulations in the oat collection, which are the position of a break point in the L(k) curve and a peak in the Δk distribution, supported values of k = 2 and k = 4 (data not shown). For both k values, most accessions were assigned by STRUCTURE to a subpopulation. With k = 2, it was possible to distinguish between the white oat landraces and the rest of the collection. With k = 4, a finer sub-grouping corresponding to the commercial varieties, the red oat landraces, and two groups of white oat landraces was obtained. Figure 2 shows the 4 subpopulations detected by STRUCTURE with the percentage of admixture of each genotype in the corresponding subpopulation. The results of the assignments showed a very good congruence between the two methods. Indeed, only 3 out of the 177 genotypes assessed, Saia, Gen51 and Gen61, were assigned to different clusters by UPGMA and STRUCTURE approaches and in the three cases STRUCTURE coefficients indicated a membership lower than 80% in the corresponding population. According to STRUCTURE, subpopulation 1 showed the lowest degree of admixture with only 16.2% of the genotypes with less than 80% of membership to this subpopulation followed by subpopulation 2 with 28.6% of genotypes with less than 80% of membership in this group. Subpopulation 3 and 4, with 31.4 and 34.6% of genotypes with less than 80% membership to the corresponding subpopulation, were the subpopulation with higher admixture. In addition the germplasm collection were also subjected to multivariate analysis based on principal component analysis (PCAs) and discriminant function analysis (DFAs) which revealed clear separation of 4 subpopulations which indicate a high consistency of the data (Figure 3).

Discussion

Genetic diversity analysis of the oat collection including landraces and cultivars revealed high variability among accessions. However, this variability was more evident within the white oat landrace collection with 129 exclusive alleles, followed by the red oat landraces, and the cultivars, with 11 and 22 exclusive alleles respectively. This indicates the considerable genetic variation that exists in landraces that is not present in the cultivars and offers opportunities for breeding new cultivars by exploiting the genetic diversity existing in the landraces. Our findings on clear distinctions between white and red oat landraces are in agreement with previous reports (Fu et al. 2005; Newell et al. 2011; Odonoughue et al. 1994) and support the hypothesis that white and red oats were domesticated independently of each other (Zohary and Hopf 2000). White oat landraces were more distinct from the white oat cultivars, however red oat landraces grouped together with the red oat cultivars analyzed (Figure 1) suggesting either lower improvement of the red oats cultivars compared with the white or the involvement of the nearest genetic similar red landraces in their genealogy.

In general, most of the accessions closely related by pedigree and/or derived from germplasm having specific traits, clustered together. Thus, those cultivars with the highest genetic similarity from cluster 1 shared common ancestors. For instance, Mirabel, Aintree, Caleche and Norlys shared Fringante as one of their parent. In addition, Norlys and Caleche shared other common ancestors. Other closely related cultivars such as AC1 and Orblanche had Mostyn as a common ancestor. Furthermore, Orblanche together with Condor, Fuwi and Adamo comprise a subcluster with a genetic

similarity of 0.7. This result may be explained by their common genealogy, since Orblanche and Adamo had Condor as an ancestor, and Adamo and Fuwi shared the cultivar Manod in their genealogy. There is less information about ancestry for the landrace collection, but geographic location of the collection sites is known (Online Resource 1). Cluster 2 grouped most red oat accessions but also four landraces described as white oats, Gen139, Gen64, Gen27 and Gen106 that also shared the 235 pb allele amplified by AM112 exclusive to this group. A detailed analysis of the morphological characters that differentiate white and red oats (Magness et al. 1971) showed that indeed Gen139, Gen64 and Gen27 in addition of the color of the seeds had the typical white oats morphological characters such as no basal scar in the spikelets from the separation from pedicel, twisted awns, and small glumes. However, Gen106 was morphologically nearer to red oat showing weak and non-twisted awns. The geographic proximity of these white oat accessions with their most related red oat landraces suggests a possible cross between the landraces so that the white landraces would have acquired some of the exclusive alleles of cluster 2. Indeed, Gen64 (white oat) and Gen65 (red oat) were sampled from the same locality while Gen106 (white oat) and Gen132 (red oat) were sampled at sites with only 8 km between them. Interestingly, the white oat landraces grouped in 2 clusters, 3 and 4. A detailed analysis of alleles showed high differences between these two clusters. For instance, more than 95 % of genotypes of cluster 3 had the alleles BarbSSR4_10-4, MAMA3-2; AM87-15 and AM04-11 whereas less than 15% of the genotypes of cluster 4 had them. By contrary more than 95% of genotypes from cluster 4 had the alleles BarbSSR_4_10-2, AME192-3 and AME192-5 whereas less than 30% of genotypes of cluster 3 had them. In order to determine the possible causes of these differences we plotted the geographic distribution of the landraces. No differences in latitude or longitude could be inferred between the

two clusters but landraces of cluster 3 were distributed in locations with significantly higher altitude than those of cluster 4 (average of 537 m altitude for cluster 3 and 377 m for cluster 4; P=0.01). This suggests a different evolution for the two clusters, with cluster 3 better adapted to higher altitudes and their associated cooler temperatures and probably poorer soils than cluster 4.

The narrow separation observed in this study among the *A. sativa* cultivars bred in European countries suggests that a rather small proportion of the available genetic variation from this species is currently used for oat improvement, as seen by Achleitner et al. (2008) in an oat collection of worldwide origin. A similar lack of diversity were also detected within sets of Canadian and Chinese oat varieties (Baohong et al. 2003; Fu et al. 2004) leading Fu and coworkers (2004) to identify an urgent need to broaden the genetic variation for sustainable oat improvement in Canada. The reduction of genetic diversity may have consequences both for the vulnerability of crops to new pests and pathogens and for their ability to respond to changes in climate and agricultural practices (Fu et al. 2003). Most of the landraces studied in this work have been characterised for disease and abiotic stress resistance (Sanchez-Martin et al. 2011a; Sanchez-Martin et al. 2011b) revealing accessions with interesting resistance that would be valuable to include in European cultivars. Study of mechanisms underlying resistance in selected landraces and varieties has been also performed (Sanchez-Martin et al. 2011a; Sanchez-Martin et al. 2011b) and the oat collection was also tested under a variety of Mediterranean environments for agronomic adaptation (unpublished results). The increased use of these accessions in European and/or Mediterranean breeding programs could simultaneously increase diversity and improve levels of valuable traits. From crosses of genetically divergent parents (i.e. a high yielding cultivar crossed with

a landrace showing disease and drought resistance) novel varieties with improved traits might be selected.

Methods such as UPGMA presented here, which do not assume predefined structure, are only loosely connected to statistical procedures allowing the identification of homogeneous clusters of individuals. For that reason, the oat germplasm collection was also analysed using an alternative model-based method implemented in the software STRUCTURE (Pritchard et al. 2000), which uses a Bayesian approach to simultaneously determine k (the number of subpopulations in a collection), and estimate for each accession the proportion of its genome that originates from each subpopulation, also called percentage of admixture. The model accounts for the presence of Hardy-Weinberg assumption or linkage disequilibrium by introducing population structure and attempts to find population groupings that (as far as possible) are not in disequilibrium (Pritchard et al. 2000).

Structure was detected in this germplasm collection using both classical multivariate and Bayesian analyses. The patterns obtained with the two methods were very similar. Population structure is the primary obstacle to successful association studies in any organism (Buckler and Thornsberry 2002). Model-based clustering suggests that a large amount of the allelic diversity can be described by subdividing the accessions into 4 discrete populations, where each subpopulation has a unique set of allele frequencies. This method is clearly a simplification of the observed data; however, it can be used to compare with other methods of clustering, and to test models of association analysis that would account for genetic associations arising from structure

presence. The congruence of patterns obtained with Bayesian and multivariate analyses suggests that the estimates of these admixture proportions are reasonably reliable.

In other cereal crops such as corn (Gunjaca et al. 2008) and rice (Bonow et al. 2009), molecular profiles associated with the description of a cultivar have been used to enforce the rights granted to breeders. The two selected markers MAMA5 and AM04 show great potential for identifying cultivars since they were able to discriminate between the 36 cultivars tested, some of them with relatively high genetic similarity, sharing common genealogy. Thus, our findings suggest that microsatellite markers can play an important role as a source of additional information in oat to supplement the morphological descriptors recommended by International Union for the Protection of New Varieties of Plants (UPOV).

Acknowledgments

This work was supported by the Spanish Ministry of Economy and Competitiveness [AGL2010-15936/AGR], the European Social and Regional Development Funds, a JAE PreDoc fellowship from CSIC to [GMB], a JAE Postdoctoral Fellowship from CSIC to [NR] and a FPU fellowship from the Spanish Ministry of Science and Innovation to [JSM]. We thank CRF (INIA, Madrid) for kindly supply the seeds of the accessions used.

References

Achleitner A, Tinker NA, Zechner E, Buerstmayr H (2008) Genetic diversity among oat varieties of worldwide origin and associations of AFLP markers with

quantitative traits. Theor Appl Genet 117 (7):1041-1053. doi:10.1007/s00122-008-0843-y

Baohong G, Zhou X, Murphy JP (2003) Genetic variation within Chinese and western cultivated oat accessions. Cereal Res Commun 31 (3-4):339-346

Becher R (2007) EST-derived microsatellites as a rich source of molecular markers for oats. Plant Breeding 126 (3):274-278. doi:10.1111/j.1439-0523.2007.01330.x

Bennett MD, Smith JB (1976) Nuclear-DNA amounts in angiosperms. Philos T Roy Soc B 274 (933):227-274

Bonow S, Von Pinho EVR, Vieira MGC, Vosman B (2009) Microsatellite markers in and around rice genes: applications in variety identification and DUS testing. Crop Sci 49 (3):880-886

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. Am J Hum Gen 32 (3):314-331

Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5 (2):107-111. doi:10.1016/s1369-5266(02)00238-8

Christiansen MJ, Andersen SB, Ortiz R (2002) Diversity changes in an intensively bred wheat germplasm during the 20(th) century. Mol Breeding 9 (1):1-11. doi:10.1023/a:1019234323372

Donini P, Law JR, Koebner RMD, Reeves JC, Cooke RJ (2000) Temporal trends in the diversity of UK wheat. Theor Appl Genet 100 (6):912-917. doi:10.1007/s001220051370

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14 (8):2611-2620. doi:10.1111/j.1365-294X.2005.02553.x

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164 (4):1567-1587

Figueiredo LFD, Sine B, Chantereau J, Mestres C, Fliedel G, Rami JF, Glaszmann JC, Deu M, Courtois B (2010) Variability of grain quality in sorghum: association with polymorphism in Sh2, Bt2, SssI, Ae1, Wx and O2. Theor Appl Genet 121 (6):1171-1185. doi:10.1007/s00122-010-1380-z

Frankel OH, Bennett E (1970) Genetic Resources in Plants—Their Exploration and Conservation. Blackwell Scientific Publications, Oxford

Fu YB, Peterson G, Scoles G, Rossnagel B, Schoen D, Richards K (2003) Allelic diversity changes in 96 Canadian oat cultivars released from 1886 to 2001. Crop Sci 43:1989-1995

Fu YB, Kibite S, Richards KW (2004) Amplified fragment length polymorphism analysis of 96 Canadian oat cultivars released between 1886 and 2001. Can J Plant Sci 84 (1):23-30

Fu YB, Peterson GW, Williams D, Richards KW, Fetch JM (2005) Patterns of AFLP variation in a core subset of cultivated hexaploid oat germplasm. Theor Appl Genet 111 (3):530-539. doi:10.1007/s00122-005-2044-2

Fu YB, Peterson GW, Chong J, Fetch T, Wang ML (2007) Microsatellite variation in *Avena sterilis* oat germplam. Theor Appl Genet 114:10229-1038

Fu YB, Williams DJ (2008) AFLP variation in 25 *Avena* species. Theor Appl Genet 117 (3):333-342

Gunjaca J, Buhinicek I, Jukic M, Sarcevic H, Vragolovic A, Kozic Z, Jambrovic A, Pejic I (2008) Discriminating maize inbred lines using molecular and DUS data. Euphytica 161:165-172

Hawkes JG (1990) The potato: evolution, biodiversity and genetic resources. Belhaven Press, London, England

He X, Bjornstad A (2012) Diversity of North European oat analyzed by SSR, AFLP and DArT markers. Theor Appl Genet 125 (1):57-70. doi:10.1007/s00122-012-1816-8

Iannucci A, Codianni P, Cattivelli L (2011) Evaluation of genotype diversity in oat germplasm and definition of ideotypes adapted to the mediterranean environment. Int J Agron 2011:8. doi:doi:10.1155/2011/870925

Jannink JL, Gardner SW (2005) Expanding the pool of PCR-based markers for oat. Crop Sci 45 (6):2383-2387. doi:10.2135/cropsci2005.0285

Jellen EN, Leggett JM (2006) Cytogenetic manipulation in oat improvement. In: Singh RJ, Jauhar JJ (eds) Genetic resources, chromosome engineering, and crop improvement: Cereals, vol 2. CRC press, pp 199-231

Karp (2002) The new genetic era: will it help us in managing genetic diversity? In: Engels JMM, Ramanatha Rao V, Brown AHD, Jackson MT (eds) Managing Plant Genetic Diversity. CAB International and IPGRI, Wallingford and Rome, pp 43-56

Koebner RMD, Donini P, Reeves JC, Cooke RJ, Law JR (2003) Temporal flux in the morphological and molecular diversity of UK barley. Theor Appl Genet 106 (3):550-558. doi:10.1007/s00122-002-1065-3

Li CD, Rossnagel BG, Scoles GJ (2000) The development of oat microsatellite markers and their use in identifying relationships among *Avena* species and oat cultivars. Theor Appl Genet 101:1259-1268

Li RZ, Wang SW, Duan LS, Li ZH, Christoffers MJ, Mengistu LW (2007) Genetic diversity of wild oat (*Avena fatua*) populations from China and the United States. Weed Sci 55 (2):95-101. doi:10.1614/ws-06-108.1

Liu ZW, Biyashev RM, Maroof MAS (1996) Development of simple sequence repeat DNA markers and their integration into a barley linkage map. Theor Appl Genet 93 (5-6):869-876. doi:10.1007/bf00224088

Magness JR, Markle GM, Compton CC (1971) Food and feed crops of the United States. In. Interregional Research Project IR-4, IR Bul. 1 (Bul. 828.New Jersey Agric. Expt. Sta.),

Murray MG, Thompson WF (1980) Rapid isolation of high molecular-weight plant DNA. Nucleic Acids Res 8 (19):4321-4325. doi:10.1093/nar/8.19.4321

Nei M, Li WH (1979) Mathematical-model for studying genetic-variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76 (10):5269-5273

Nersting LG, Andersen SB, von Bothmer R, Gullord M, Jorgensen RB (2006) Morphological and molecular diversity of Nordic oat through one hundred years of breeding. Euphytica 150 (3):327-337. doi:10.1007/s10681-006-9116-5

Newell MA, Cook D, Tinker NA, Jannink JL (2011) Population structure and linkage disequilibrium in oat (*Avena sativa* L.): implications for genome-wide association studies. Theor Appl Genet 122 (3):623-632. doi:10.1007/s00122-010-1474-7

Odonoughue LS, Souza E, Tanksley SD, Sorrells ME (1994) Relationships among North-American oat cultivars based on restriction-fragment-length-polymorphisms. Crop Sci 34 (5):1251-1258

Paczos-Grzeda E (2004) Pedigree, RAPD and simplified AFLP-based assessment of genetic relationships among *Avena sativa* L. cultivars. Euphytica 138 (1):13-22. doi:10.1023/B:EUPH.0000047055.99322.7a

Pal N, Sandhu JS, Domier LL, Kolb FL (2002) Development and characterization of microsatellite and RFLP-derived PCR markers in oat. Crop Sci 42 (3):912-918

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155 (2):945-959

Pritchard JK, Wen W (2003) Documentation for STRUCTURE software: Version 2. . Available from http://pritchbsduchicagoedu

Reif JC, Zhang P, Dreisigacker S, Warburton ML, van Ginkel M, Hoisington D, Bohn M, Melchinger AE (2005) Wheat genetic diversity trends during domestication and breeding. Theor Appl Genet 110 (5):859-864. doi:10.1007/s00122-004-1881-8

Roussel V, Koenig J, Beckert M, Balfourier F (2004) Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes. Theor Appl Genet 108 (5):920-930. doi:10.1007/s00122-003-1502-y

Sanchez-Martin J, Rubiales D, Prats E (2011a) Resistance to powdery mildew (*Blumeria graminis* f.sp *avenae*) in oat seedlings and adult plants. Plant Pathol 60 (5):846-856. doi:10.1111/j.1365-3059.2011.02453.x

Sanchez-Martin J, Rubiales D, Sillero JC, Prats E (2011b) Identification and characterization of sources of resistance in *Avena sativa*, *A. byzantina* and *A. strigosa* germplasm against a pathotype of *Puccinia coronata* f.sp. *avenae* with

virulence against the Pc94 resistance gene. Plant Pathol. doi:10.1111/j.1365-3059.2011.02514.x

Stevens EJ, Armstrong KW, Bezar HJ, Griffin WB, J.G. H (2004) Fodder oats an overview. In: Suttie JM, Reynolds SG (eds) Fodder oats: A world overview. Food and Agriculture Organization of the United Nations, Rome, pp pp. 1-9

Tessier C, David J, This P, Boursiquot JM, Charrier A (1999) Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. Theor Appl Genet 98 (1):171-177. doi:10.1007/s001220051054

Warburton ML, Peif JC, Frisch M, Bohn M, Bedoya C, Xia XC, Crossa J, Franco J, Hoisington D, Pixley K, Taba S, Melchinger AE (2008) Genetic diversity in CIMMYT nontemperate maize germplasm: Landraces, open pollinated varieties, and inbred lines. Crop Sci 48 (2):617-624. doi:10.2135/cropsci2007.02.0103

Wight CP, Yan WK, Fetch JM, Deyl J, Tinker NA (2010) A set of new simple sequence repeat and *Avenin* DNA markers suitable for mapping and fingerprinting studies in oat (*Avena* spp.). Crop Sci 50 (4):1207-1218. doi:10.2135/cropsci2009.09.0474

Zohary D, Hopf M (2000) Domestication of plants in the old world. Third edn. Oxford University Press, Oxford-New York

Table 1: Characteristics of 31 microsatellite markers selected for use in the study**.** Size range, number of alleles (rare, common and most frequent), confusion probability (Cj), discriminating power (Dj) and polymorphic information content (PIC).

Marker name	SSR type	Tm	Detection[c]	Size (pb)	N° alleles	Rare alleles (<0,03)	Common alleles (0,03-0.2)	Frequent alleles (>0,2)	*Cj*	*Dj*	PIC
AM01	Genomic	55	ABI	154-240	27	10	14	3	0.11	0.89	0.89
AM03	Genomic	58	ABI	249-298	22	7	15	0	0.06	0.94	0.93
AM04	Genomic		ABI	78-180	41	7	28	6	0.04	0.96	0.96
AM07	Genomic	55	ABI	146-195	25	8	13	4	0.08	0.92	0.92
AM102	Genomic	55	ABI	160-217	10	2	3	5	0.18	0.82	0.82
AM112	Genomic		ABI	227-255	7	3	3	1	0.54	0.46	0.46
AM14	Genomic	55	ABI	98-134	17	4	5	8	0.10	0.90	0.90
AM30	Genomic	55	ABI	178-230	17	3	5	9	0.11	0.89	0.89
AM42	Genomic	58	ABI	165-208	13	4	1	8	0.14	0.86	0.86
AM87	Genomic	55	ABI	92-171	15	6	3	6	0.17	0.83	0.83
AM89	Genomic	53	ABI	173-201	10	3	2	5	0.22	0.78	0.77
AME097	EST	52	Silver	145-155	4	0	0	4	0.26	0.74	0.74
AME105	EST	52	Silver	140-190	10	0	5	5	0.13	0.87	0.87
AME168	EST	52	Silver	200-220	3	0	1	2	0.54	0.46	0.46
AME176	EST	52	Silver	90-110	4	0	1	3	0.32	0.68	0.68
AME192	EST	52	Silver	300-345	6	1	0	5	0.22	0.78	0.78
BarbSSR_2-40	EST	60	Silver	195-220	4	0	0	4	0.37	0.63	0.63
BarbSSR_4-10	EST	60	Silver	270-310	5	1	1	3	0.36	0.64	0.63
CDO187	EST	55	ABI	104-152	9	3	0	6	0.18	0.82	0.82
Fesc12	EST	61	ABI	124-194	21	7	12	2	0.10	0.90	0.90
HVM20	Genomic	53	ABI	103-154	21	9	11	1	0.10	0.90	0.90
HvXan	EST	50	ABI	93-206	26	15	6	5	0.15	0.85	0.85
MAMA01	Genomic	55	ABI	183-215	12	3	7	2	0.18	0.82	0.81
MAMA03	Genomic	55	ABI	351-403	12	2	8	2	0.14	0.86	0.85
MAMA05	Genomic	55	ABI	62-274	42	16	20	6	0.07	0.93	0.92
MAMA07	Genomic	55	ABI	322-371	12	5	5	2	0.29	0.71	0.70
MAMA08	Genomic	55	ABI	548-623	18	5	12	1	0.10	0.90	0.90
MAMA09	Genomic	55	ABI	401-491	14	6	4	4	0.16	0.84	0.84
MAMA11	Genomic	55	ABI	124-183	9	4	3	2	0.24	0.76	0.75
MAMA12	Genomic	55	ABI	297-321	8	2	3	3	0.22	0.78	0.78
OL0410	EST	55	ABI	256-281	10	2	3	5	0.19	0.81	0.81
Mean					**14.65**	**4.45**	**6.26**	**3.94**	**0.20**	**0.80**	**0.80**
Total					454	138	194	122			

[c]ABI; Applied Biosystem ABI3137

Table 2: Comparison of allelic frequency between subpopulations.

	Alellic Frecuency				Signific
PRIMER	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	
AM01	0,12[a]	0,11[a]	0,11[a]	0,12[a]	ns
AM03	0,05[ac]	0,06[ad]	0,07[bd]	0,05[c]	***
AM04	0,12[a]	0,13[a]	0,13[a]	0,10[b]	***
AM07	0,09[ab]	0,10[a]	0,08[b]	0,10[a]	ns
AM102	0,21[a]	0,31[b]	0,23[a]	0,16[c]	***
AM112	0,18[a]	0,18[a]	0,15[bc]	0,15[c]	***
AM14	0,21[a]	0,20[a]	0,25[b]	0,22[ab]	*
AM30	0,33[a]	0,43[b]	0,38[c]	0,33[a]	***
AM42	0,32[a]	0,31[a]	0,36[b]	0,31[a]	**
AM87	0,14[a]	0,16[a]	0,18[b]	0,21[c]	***
AM89	0,33[a]	0,34[a]	0,33[a]	0,34[a]	ns
AME097	0,36[ab]	0,43[a]	0,34[b]	0,33[b]	***
AME105	0,14[a]	0,26[b]	0,24[b]	0,24[b]	***
AME168	0,39[a]	0,44[a]	0,41[a]	0,53[b]	***
AME176	0,48[a]	0,48[a]	0,63[b]	0,59[b]	***
AME192	0,36[a]	0,43[b]	0,43[b]	0,38[a]	**
BarbSSR_2-40	0,32[a]	0,49[b]	0,45[ab]	0,52[b]	*
BarbSSR_4-10	0,40[a]	0,45[b]	0,43[ab]	0,41[a]	***
CDO187	0,38[a]	0,43[bc]	0,46[b]	0,40[ac]	**
FESC12	0,09[a]	0,09[a]	0,08[a]	0,09[a]	ns
HVM20	0,06[a]	0,07[a]	0,08[a]	0,08[a]	ns
HvXan	0,16[a]	0,15[a]	0,17[ac]	0,19[bc]	*
MAMA01	0,10[a]	0,09[ac]	0,08[bc]	0,10[a]	*
MAMA03	0,10[a]	0,11[a]	0,10[a]	0,16[b]	***
MAMA05	0,09[a]	0,10[ab]	0,11[b]	0,08[c]	***
MAMA07	0,11[ab]	0,12[a]	0,11[ab]	0,10[b]	ns
MAMA08	0,07[ac]	0,08[ab]	0,06[c]	0,06[c]	ns
MAMA09	0,09[a]	0,13[b]	0,14[b]	0,12[b]	***
MAMA11	0,14[a]	0,11[b]	0,11[b]	0,12[b]	***
MAMA12	0,14[a]	0,17[a]	0,16[a]	0,22[b]	***
OL0410	0,32[a]	0,35[a]	0,40[b]	0,34[a]	***

Different letter in a row indicates significant differences at $p \leq 0.05$, $p < 0.01$ and $p \leq 0.001$ according to *, **, and *** respectively for that marker

Figure Captions

Fig. 1 UPGMA dendogram of 176 oat accessions based on DICE distance for 31 SSR markers

Fig. 2 Estimated population structure of oat genotypes according to STRUCTURE software. Each individual is represented by a thin vertical segment, which can be partitioned into 4 grey-scale colored segments that represent the individual estimated membership to the 4 clusters

Fig. 3 Scatterplot of Discriminant Function Analysis scores of components 1 and 2 based on 31 SSR markers used in this study. A. Represented are the genotypes belonging to cluster 1(circles), cluster 2 (squares), cluster 3 (crosses) and cluster 4 (triangles). B. Magnification of the central part of the graph A with the groups assigned according to the DFA scores

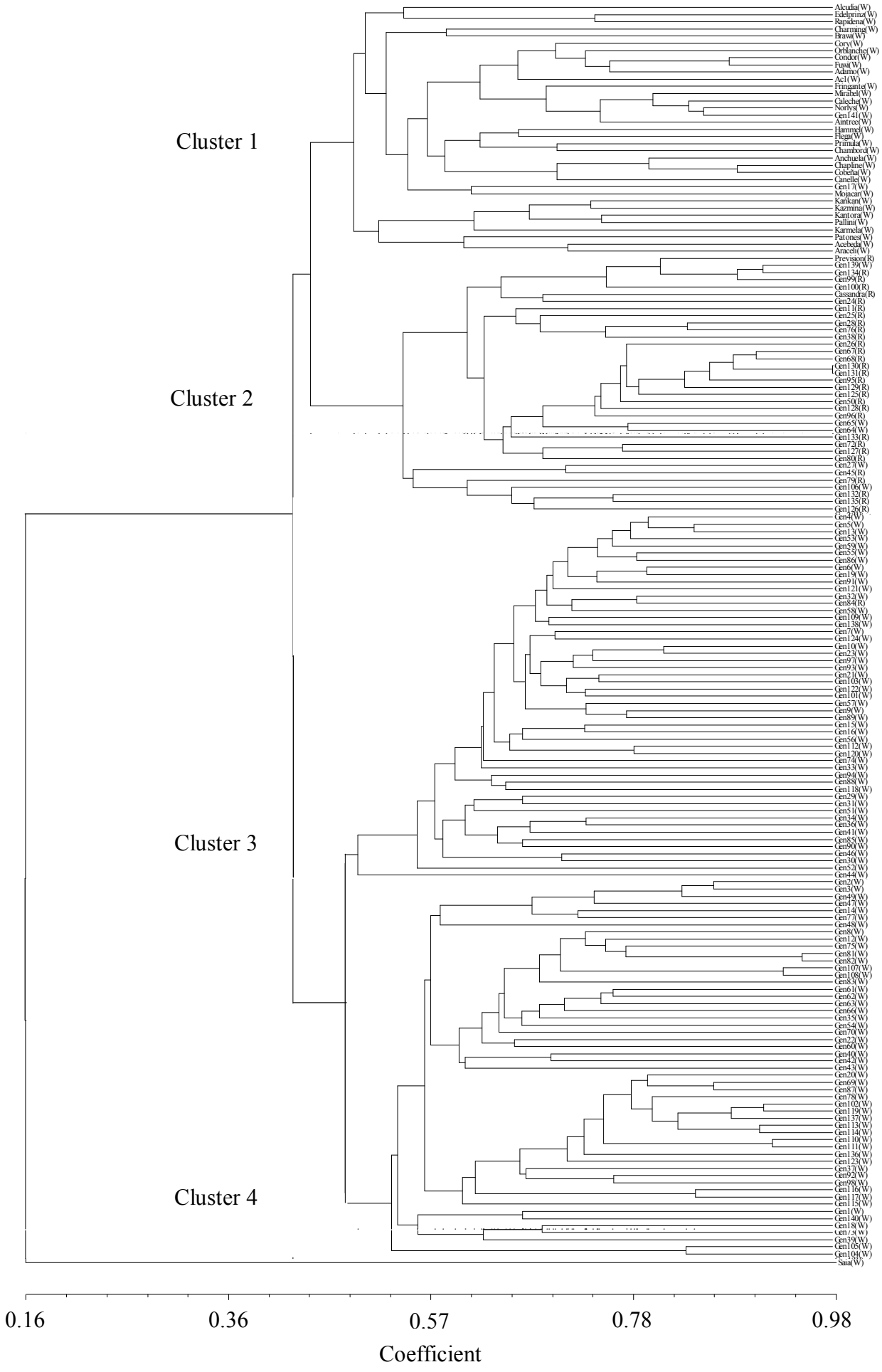
Cluster 1
Cluster 2
Cluster 3
Cluster 4
0.16
0.36
0.57
0.78
0.98
Coefficient

Figure 1.

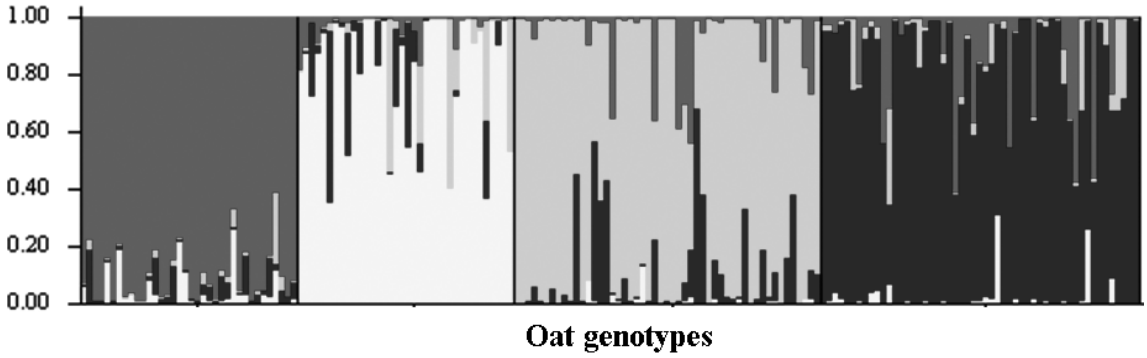

Figure 2.

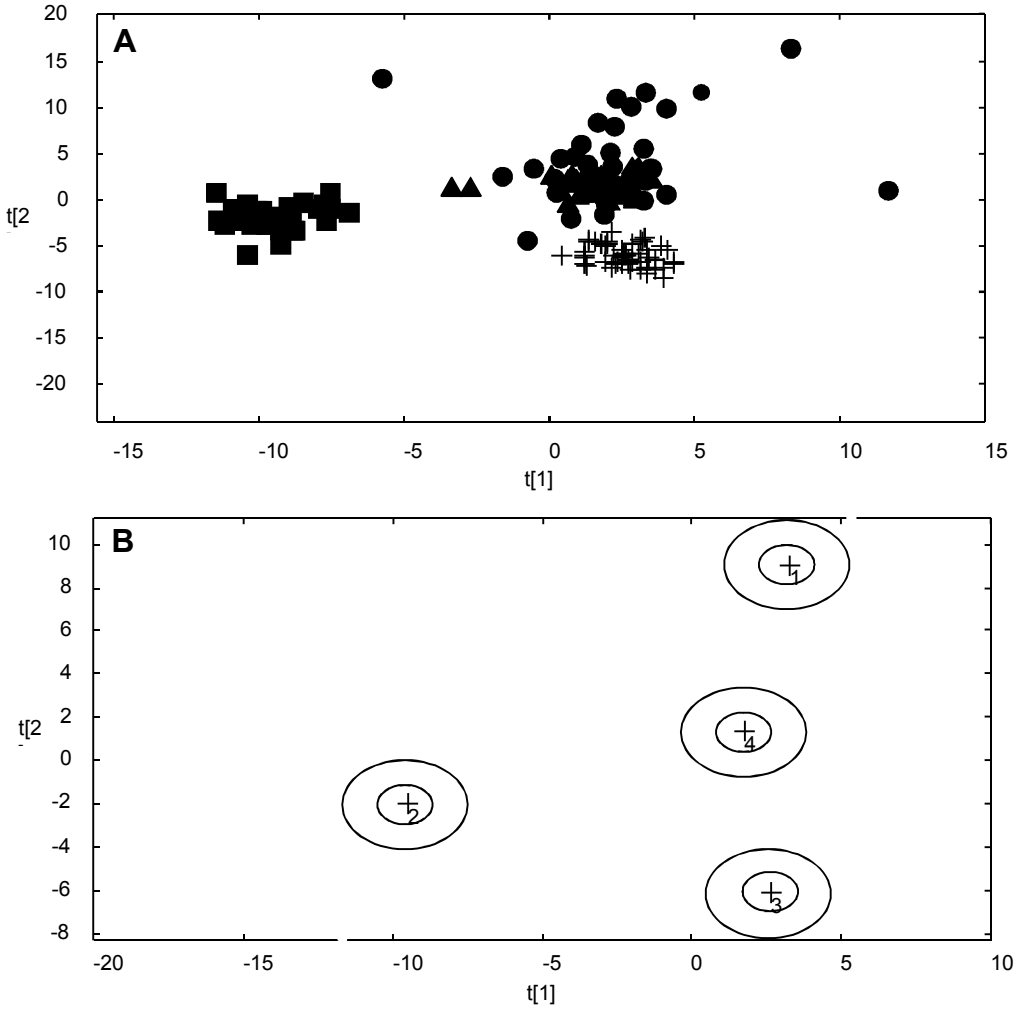

Figure 3.